Temporal Constraints on the Incorporation of Regulatory Mutants in Evolutionary Pathways

Temporal Constraints on the Incorporation of Regulatory Mutants in Evolutionary Pathways

Kyle M. Brown[1]*, Mark A. DePristo[1,2], Daniel M. Weinreich[1,3], Daniel L. Hartl[1]

Running Title:

[1]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138

[2]Present Address:
Broad Institute, 7 Cambridge Center, Cambridge, MA 02142

[3] Present Address:
Department of Ecology and Evolutionary Biology and Center for Computational, Molecular Biology, Brown University Box G-W, Providence, RI 02192

*Corresponding Author:
Kyle M. Brown, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138
Phone: 617-496-5540
Fax: 617-496-5854
kylemichaelbrown@gmail.com

Text pages:

Figures:

Tables:

Abstract

Understanding the molecular details of the sequence of events in multistep evolutionary pathways can reveal the extent to which natural selection exploits regulatory mutations affecting expression, amino acid replacements affecting the active site, amino acid replacements affecting protein folding or stability, or variations affecting gene copy number. In experimentally exploring the adaptive landscape of the evolution of resistance to β-lactam antibiotics in enteric bacteria, we noted that a regulatory mutation that increases β-lactamase expression by about twofold has a very strong tendency to be fixed at or near the end of the evolutionary pathway. This pattern contrasts with previous experiments selecting for the utilization of novel substrates, in which regulatory mutations that increase expression are often fixed early in the process. To understand the basis of the difference, we carried out experiments in which the expression of β-lactamase was under the control of a tunable arabinose promoter. We find that the fitness effect of an increase in gene expression is highly dependent on the catalytic activity of the coding sequence. An increase in expression of an inefficient enzyme has a negligible effect on drug resistance, however the effect of an increase in expression of an efficient enzyme is very large. The contrast in the temporal incorporation of regulatory mutants between antibiotic resistance and the utilization of novel substrates is related to the nature of the function that relates enzyme activity to fitness. A mathematical model of β-lactam resistance is examined in detail, and shown to be consistent with the observed results.

Introduction

Much discussion has focused on the relative role of structural versus regulatory mutations in the evolution of novel phenotypes. Structural changes include amino acid replacements (e.g., Clark *et al.* 2003, Hoekstra *et al.* 2006), and regulatory mutations include those that alter gene expression in *cis* or in *trans* (e.g., Olds and Sibley 2003, Shapiro *et al*. 2004, 2006, Tishkoff *et al.* 2007, Brown *et al.* 2008). Various perspectives are summarized in Carroll (2000, 2005a, 2005b), Wray (2007), Hoekstra and Coyne (2007), and Lynch and Wagner (2008).

In this paper, we take a different tack. We consider the evolution of metabolic capabilities to which both structural and regulatory mutations are likely to contribute. We ask why it is that, in some systems, regulatory mutations are incorporated early in the process; whereas, in other systems, regulatory mutations are incorporated late.

Extensive previous research has observed that regulatory mutations often precede structural ones in enzyme evolution (Mortlock *et al.* 1965, Wu *et al.* 1968, Hegeman and Rosenberg 1970, Hall and Hauer 1993, among others). In these situations, existing enzymes often catabolize novel substrates to some extent, but they require constitutive regulatory mutations in order allow sufficient expression to enable growth. Similarly, cryptic genes for the metabolism of certain substrates reside unexpressed in microbial genomes until mutationally activated by promoter mutations (Hall *et al*. 1983, Hall 1998). As the initial substitution in an adaptive landscape is predicted to account for ~30% of the total fitness increase (Orr 2002), these observations suggest that regulatory mutations play a key role in enzyme evolution.

"Regulation first" has some notable exceptions, however. For example, Weinreich *et al.* (2006) found that structural mutations usually precede regulatory mutations in the

evolution of the TEM β–lactamase in *E. coli*. In the adaptive landscape connecting the wildtype TEM allele of low resistance to a quintuple mutant of high resistance, Weinreich *et al.* (2006) showed that a particular regulatory mutation denoted g4205a has a 75% chance of being the final mutation fixed. Similarly, studies on an evolved β–galactosidase enzyme derived from the *E. coli* gene *ebg* have shown that this enzyme requires an initial structural mutation in order to facilitate growth on its substrate (Hartl & Hall 1974, Hall & Hartl 1974, Hall 1990).

In this paper, we show how temporal constraints on the incorporation of regulatory mutations are associated with the catalytic activity of the genes involved, and with the differing relationships between enzyme activity and fitness for metabolic and antibiotic-resistance enzymes. By means of studies of resistance to the β–lactam antibiotic cefotaxime in strains of *E. coli* containing the TEM β–lactamase, we empirically demonstrate the importance of structural mutations incorporated early in the evolutionary pathway of drug resistance. We also show that the temporal ordering of structural versus regulatory mutations in evolution depends on the mapping of enzyme activity onto fitness.

Results

To sharpen the discussion, Figure 1 depicts three contrasting functions relating fitness to enzyme activity. The concave function (dashed line) depicts a relationship common for many metabolic enzymes, and the mapping is appropriate when metabolic flux serves as a proxy for fitness (Hartl *et al.* 1985, Dykhuizen *et al.* 1987). The convex function (dotted line) is common to enzyme-mediated antibiotic resistance. While this specific model has been used to successfully predict β–lactam resistance in several bacterial species (Zimmerman and Rosselet 1977, Nikaido and Normark 1987, Lakaye *et al.* 1999), its implications for the temporal incorporation of structural versus regulatory mutations has not been explored.

In the present studies, we constructed all combinations of three TEM β–lactamase mutations associated with increased resistance and placed them under the control of an inducible and titratable promoter derived from the arabinose operon (Materials and Methods). The rationale is that β–lactam resistance is affected by both structural mutations via changes in apparent affinity (k_{cat}/K_M) and by regulatory mutations mediated by changes in promoter sequences altering gene expression and therefore enzyme concentration (Zimmerman and Rosselet 1977). For each of the eight TEM β–lactamase alleles, we measured resistance across a range of expression levels (Material and Methods). Resistance was assayed as the minimal inhibitory concentration (MIC), the smallest concentration of cefotaxime that completely inhibits growth. For the kinetic parameters of these enzymes toward cefotaxime, we used previously published data (Wang *et al.* 2002).

The key discovery was that the effect of increased expression on drug resistance was highly dependent on the TEM structural gene. Most striking, alleles that contain the mutation Gly238Ser (G238S) result in large increases in resistance with increased expression, whereas alleles retaining the ancestral Gly at position 238 show no more than a twofold increase in resistance across a more than 100-fold increase in transcription (Supplementary Figure S1). While the sequence-dependent effect of increasing gene expression is most dramatic for the mutation G238S, the mutations Glu104Lys (E104K) and Met182Thr (M182T) also show modest effects (Supplementary Figure S2).

To quantify the effect of structural mutations on the fitness effects of increased expression, we developed a generalized linear model (Materials and Methods) of antibiotic resistance (MIC) as a function of both coding sequence and expression level. Among the 64 MIC's in our dataset, we find significant effects attributable to the independent contribution of each individual mutation (G238S, E104K, M182T) as well as expression level (*F* test, *P*-values: G238S = 2.2×10^{-16}, E104K = 1.914×10^{-12}, M182T = 0.00035, expression level = 1.548 10^{-15}). There is, however, a highly significant interaction between expression level and G238S (*F* test, Expression × G238S *P*-value = 4.776×10^{-6}). Although interactions between regulatory and coding mutations have been noted previously (Stam & Laurie 1996, Weinreich *et al.* 2006), our experimental design allows a formal statistical confirmation of Expression × G238S epistasis.

To investigate in greater detail how inducer (arabinose) concentration corresponds to expression level, we measured mRNA and protein concentration for a subset of alleles across a range of induction levels. Alleles with different coding mutations had similar mRNA levels at the same level of induction (Supplementary Figure S1). However, some

alleles result in different steady-state enzyme concentrations at the same concentration of inducer (Supplementary Figure S3, Table S4). These results are consistent with previous work demonstrating differences in stability among these proteins *in vitro* (Wang *et al.* 2002). The ranges in allele-specific protein abundance we observe are also consistent with other observations relating mRNA level to steady-state protein abundance (Ghaemmaghami *et al*. 2003).

Our data on protein abundance support the hypothesis that the structural mutation G238S is the one principally responsible for the temporal phasing of the regulatory mutation g4205a in the evolution of TEM β–lactamase. Correlation analysis between protein abundance and resistance demonstrate that resistance is significantly correlated with relative TEM β-lactamase abundance for alleles containing G238S (G238S and M182T+G238S; Spearman's rank correlation: $P = 0.0005641$). However, we find no such correlation for the alleles containing a G at site 238 (wild type and M182T alleles; Spearman's rank correlation: $P = 0.287$).

As a final step in the analysis, we used known kinetic parameters and our measured protein abundances to test whether a previously described fitness function for TEM β-lactamase (Zimmermann and Rosselet 1977) predicted the observed relationship between structural mutations and expression level. Figure 3 shows that the predicted resistance values (solid lines) do match the observed resistance values (open circles) across the range of protein abundance. This result suggests that the Zimmermann-Rosselet (1977) model is a good predictor of antibiotic resistance. Further exploration of the evolutionary implications of this model might yield additional insights.

Discussion

Our results explain the temporal phasing of the incorporation of the regulatory mutation g4205a in the evolutionary pathway of TEM β-lactamase. The most important constraint is the identity of the residue at amino acid site 238. There is a highly significant epistatic interaction between G238S and expression level. In the presence of wildtype Gly238, the g4205a mutation yields little or no increase in antibiotic resistance, whereas in the presence of G238S, the effect can be large (Weinreich *et al.* 2006).

The G238S mutation has such a dramatic effect because of its impact on the β-lactamase's affinity for its substrate, cefotaxime. The G238S substitution alone increases the overall rate of hydrolysis of cefotaxime by more than 65-fold (wildtype reaction rate constant (k_{cat}) = 0.636 s^{-1}; G238S reaction rate constant (k_{cat}) = 41.8 s^{-1} (Wang *et al*. 2002). This change in reaction rate (k_{cat}), along with a smaller decrease in the Michaelis constant (K_M), creates a nearly 100-fold higher apparent affinity for cefotaxime [wildtype: $k_{cat}/K_M = 2.07 \times 10^3$ s^{-1} M^{-1}, G238S: $k_{cat}/K_M = 1.78 \times 10^5$ s^{-1} M^{-1}] (Wang et al. 2002).

Increased substrate affinity allows for lower concentrations of enzyme to produce the same *in vivo* activity. The Zimmerman-Rosselet (1977) model of resistance predicts, and our data confirm, that at the same *in vivo* concentrations of β-lactamase, alleles with G238S have increased resistance while those without it do not (Supplementary Figure S4, Figure 3).

In other words, G238S alters the mapping of enzyme activity onto fitness. The effect is explored in Figure 4, which compares the effects of increasing substrate affinity in concave (top) versus convex (bottom) fitness regimes. Each line represents the effect

of increasing enzyme concentration for an enzyme with a given set of kinetic parameters. Both sets of curves exhibit fitness plateaus, where increases in enzyme concentration have little or no effect on fitness. In these regions of the curves, structural mutations are likely to be most important because structural mutations can increase fitness by allowing higher activity from the same concentration of enzyme. These kinds of changes can allow jumps from one fitness curve to another. In this way structural mutations that alter substrate affinity can change the rules of the game. On the other hand, in regions of the curve where fitness increases steeply with activity, regulatory mutations that increase enzyme concentration can be strongly selected.

Figure 4 reconciles the contrasting tempos of when regulatory mutants are likely to be incorporated into evolutionary pathways. In the evolution of flux-limited metabolic pathways illustrated by the concave fitness curves, mutations that increase gene expression are likely to be incorporated early, since an increase in expression can have a large effect on fitness. In the evolution of antibiotic resistance, illustrated by the convex fitness curves, mutations that increase gene expression are likely to be incorporated later, since increased expression of catalytically inefficient enzymes contributed negligibly to fitness.

Materials and Methods

Construction of E. coli *Strains*

The TEM-1 gene from the plasmid pBR322 was isolated via PCR and cloned into the pBAD vector with a kanamycin resistance marker using the pBAD TOPO TA expression kit (Invitrogen, Carlsbad, CA). Point mutations were introduced into this gene via site-directed mutagenesis according to the Quick-change site directed mutagenesis kit (Stratagene, La Jolla, CA). The entire TEM locus and pBAD promoter was then sequenced to verify the presence of the desired point mutations and the absence of all other substitutions.

In order to ensure that we could modulate the transcript levels in individual cells, and not merely the population as a whole, we obtained the K12-derived cell line, BW27783 (Khlebnikov *et al.* 2001; J. Keasling, Univ. California, Berkeley) and transformed our pBAD plasmids bearing the TEM alleles into this strain. Previous work had demonstrated that the arabinose operon behaved in individual cells in an "all on" or "all off" fashion (Khlebnikov *et al.* 2001). The BW27783 cell line has the promoter of its araE gene replaced with a constituitive promoter, eliminating positive feedback of this expression system. Using the BW27783 cell line, arabinose concentrations should correlate to the RNA abundance of the arabinose operon-controlled TEM locus in each cell.

Resistance Assays

We used minimum inhibitory concentration (MIC) assays to measure resistance to the β-lactam antibiotic cefotaxime (Sigma-Alrich, St. Louis, MO). This method is

detailed by the Clinical and Laboratory Standards Institute (Performance standards… 2007). Briefly, two fold dilution series over an appropriate range (0.03125–1024 ug/ml) of cefotaxime concentrations in Mueller-Hinton broth were prepared in 96 well flat bottom plates. Strains were struck out on fresh Luria Broth (LB) and kanamycin (KAN; 50 ug/ml) plates. Colonies were picked the following day and grown overnight to saturation in LB-KAN media containing the appropriate arabinose concentration. Ninety-six well MIC assay plates were then inoculated with ~10^5 cells/ml in each well as determined by cell titre counts on LB-KAN plates. MIC values from wells inoculated with between 10^4 and 10^6 cells/liter were insensitive to differences in cell concentration. Wells inoculated with concentrations higher than 10^6 or lower than 10^4 cells/ml demonstrated dramatically increased and decreased MIC readings respectively and were not used. For each overnight culture of a given strain and arabinose concentration, MIC assays were done in triplicate and the median MIC was determined from these three values.

Transcription Induction Confirmation

Quantitative reverse transcriptase PCR (Q-rtPCR) reactions were used to confirm high levels of TEM allele transcription from pBAD plasmids in the presence of high concentrations to arabinose. Two or three biological replicates for each combination of strains and arabinose concentration were prepared by inoculating 1ml overnight cultures. After approximately 18 hours of growth, the cultures were diluted 10,000 times and grown to log phase (OD_{600} ~ 0.6). One RNA extraction per biological replicate was performed using the RNeasy kit (Qiagen, Valencia, CA) and treated with Turbo-

DNAFree (Ambion, Austin, TX). Adding the same quantity of RNA to each reaction, we created cDNA from total RNA using Random Hexamer Primers and Superscript II (Invitrogen, Carlsbad, CA). As previous experiments had shown there to be little variation between cDNA reactions from the same RNA preparation, we prepared only one cDNA reaction per RNA sample. Quantitative PCR (QPCR) reactions using SYBR green (Qiagen, Valencia, CA) were used to quantify cDNA concentrations in each RNA sample. At least six QPCR reactions (3 experimental and 3 control) were run in parallel for each RNA sample. Primers for both the TEM alleles and the control tRNA gene *trpT* (EG30105) were designed to amplify with 95% or greater efficiency, calculated based on the standard curve for each (see *Statistical Analysis* below). Uninduced and induced cells demonstrated significantly different expression levels. The wildtype and G238S+M182T alleles demonstrated similar mRNA levels at the same arabinose concentrations (Figure S1).

Protein Purification

In order to quantify the TEM enzymatic concentrations in our cells at different arabinose concentrations, we extracted and purified four of the β-lactamase missense mutants (all combinations of M182T and G238S, see Table S1). Briefly, each allele was moved into a pBAD-based arabinose-inducible over-expression and purification vector (pBAD Directional TOPO expression kit, Invitogen) carrying a 28 residue C-terminal linker including a 6xHis tag. The alleles were constructed with a combination of DNA digestion with the NcoI endonuclease (New England Biolabs) and ligation (EMD

Biosciences) or site-directed mutagenesis (Weinreich *et al.* 2006). All sequences were verified by DNA sequencing of purified plasmids.

Overnight cultures were grown in Terrific Broth containing 0.1% arabinose for each of the missense alleles in *E. coli* strain LMG194 (Invitrogen). Purified β-lactamase was obtained from 5 ml cultures using His-select iLAP columns (Sigma-Aldrich). Purity of the resulting elutions was verified using SDS-PAGE. Pure protein concentrations were between 10-1000 μg/ml as determined by Bradford assays (Quick Start Bradford Protein Assay, Bio-Rad, Hercules CA). The purified proteins were stored at -80ºC in 40 mM $NaPO_4$ containing 40% glycerol. Note that neither imidazole nor elution salts were removed from the purified proteins due to problems with protein aggregation and precipitation from the unstable alleles. Neither additive had a significant or systematic effect on kinetics assays performed in this study.

Enzyme Kinetics

Enzyme reaction kinetics were determined using purified protein against nitrocefin (NTF; Oxoid Ltd, Basingstoke, Hampshire, England; λ=486 nm, Δε=15,000 M^{-1} cm^{-1}, concentrations from 10-200 μM) at 25ºC in 100 mM $NaPO_4$, pH 7 buffer containing 2000 μg/ml bovine serum albumin (Sigma #A3059-50G) as an enzyme stabilizer (Laraki *et al.* 1999) on a 96-well spectrophotometer (Spectramax PLUS^{384} from Molecular Devices). Initial reaction velocities were fit to the complete Michaelis-Menten equation (Equation 1) using non-linear least-squares for high-performance alleles (those with K_M's less than half the maximum substrate concentration) or to the reduced Michaelis-Menten equation where $S << K_M$ using standard linear regression (Equation 2)

where K_M is the Michaelis constant, k_{cat} is the general rate constant, and S is the concentration of substrate, E_0 is the enzyme concentration and v the reaction rate (Nelson and Cox 2000).

$$v = \frac{E_0 \cdot k_{cat} \cdot S}{K_M + S} \quad (1)$$

$$v = E_0 \cdot \frac{k_{cat}}{K_M} \cdot S \quad (2)$$

All fits had P-values < 10^{-3}. Presented values are the average of six replicates. Standard errors of the mean are less than 4% of the average value (Table S1). Individual k_{cat} and K_M for each allele against NTF are presented in Table S1.

Protein Extractions and Enzyme Concentrations

We quantified the enzyme concentration in four of our TEM-allele expressing BW27783 cell lines at six of the eight arabinose concentrations used in MIC experiments. For each biological replicate of an allele at a given arabinose concentration, a separate colony was picked into liquid LB-Kan containing the appropriate arabinose concentration and grown for 24 hours. Following a method similar to that for MIC assays, overnight liquid cultures were used to inoculate ninety-six well MIC assay plates with 200 μl of MH broth containing arabinose at a cell concentration of ~10^5 cells/ml. These cultures were grown without shaking for 24 hours. We extracted total soluble protein from the combined volume of eight wells (~1.5ml of cell culture) using the B-PER® Bacterial Protein Extraction Reagent (Pierce, Rockford, IL) according to manufacturers instructions. We performed three biological replicates for each allele and arabinose concentration tested.

Using the kinetic parameters we calculated for individual β-lactamase alleles toward nitrocefin (NTF) (Table S1; see above), we used dilute concentrations of protein extractions in saturating concentrations (50 μM) of nitrocefin to determine enzyme concentration in each protein extraction. Under these conditions, the rate of color change of the NTF solution is equal to the maximum rate of enzymatic catalysis, referred to as the V_{max}. For each protein extraction, we measured V_{max} by measuring absorbance at 486nm every 10 seconds for 5-10 minutes. Only protein extraction dilutions that yielded linear curves over the entire time course were used to measure V_{max} (milliΔOD/min). We performed four technical replicates for each protein extraction.

After normalizing by the amount of total soluble protein extract (μg) used (Bradford assay, see above), we substituted our calculated V_{max} for v in the inverted Michaelis-Menten equation (Equation 1) to calculate the enzyme concentration in each protein extraction. These calculations yielded protein concentrations in the unconventional units of milliΔOD/60μg total soluble extract. After determining the protein concentration in four technical replicates for each biological replicate, we calculated means and standard errors across the three biological replicates (See Supplementary Figure S3).

Statistical Analysis

Based on the maximum-likelihood approach for reporting MIC values described previously (Weinreich *et al.* (2006), we report the median MIC of the three replicate assays performed for each allele at each arabinose concentration.

Relative RNA expression levels based on Quantitative PCR (Q-PCR) reactions of cDNA created from total RNA extractions of wildtype and G238S+M182T alleles at 0% and 10^{-2}% arabinose were statistically analyzed using the following modified ΔCT approach. Briefly, Q-PCR reactions produced raw CT values for TEM β-lactamase and *trpT* tRNA for each cDNA sample. The replicate CT values for the *trpT* gene for each c-DNA sample were averaged. This control CT value was then subtracted from each replicate CT value for β-lactamase to yield ΔCT. A nested analysis of variance revealed much greater variation in ΔCT values among biological replicate than within them. We averaged the replicate ΔCT values obtained from each biological replicate (cDNA preparation). Heteroscedasticity *t*-tests were used to test for the homogeneity among mean ΔCT values. Analysis of variance calculations and *t*-tests were carried out in R (version 2.5.0; www.R-project.org) and additional *t*-tests were carried out using in Microsoft Excel 2004 for Mac (version 11.3.7).

Relative RNA expression levels reported in Figure S1 were calculated as follows. Following typical quantitative-PCR methodology involving a standard curve, a genomic DNA extraction from cells harboring a TEM-1 containing plasmid (pBR322) was used to create a 10 fold DNA dilution series over 6 orders of magnitude. CT values from these samples for each primer set were then plotted on a log-log scale against their relative concentrations. Using the best-fit line determined by these points, we used the mean absolute CT value for each cDNA to determine the relative concentration of cDNA of either TEM β-lactamase or *trpT* for each biological replicate. We normalized the TEM β-lactamase transcription levels by dividing by the *trpT* transcription levels calculated for each cDNA. The calculated normalized relative quantities of TEM β-lactamase were then

averaged for each genotype by arabinose combination and plotted as mean relative RNA concentrations in Figure S1. The error bars in Figure S1 represent the standard error among the normalized relative quantities calculated for each biological replicate.

We constructed a generalized linear model (GLM, function glm() in *R*, ver. 2.2.1, The R Foundation for Statistical Computing) to analyze our MIC data. We modeled MIC as a function of the presence or absence of the three mutations (G238S, E104K, M182T) and expression level as determined by inducer concentration. For our GLM analysis, we gave each median MIC value in our data set an ordinal number where 1 corresponded to the lowest expression level (growth in 0% arabinose) and 8 corresponded to the highest expression level (growth in 10^{-1}% arabinose). As our MIC measurements represent quantized values and appeared underdispersed as a Poisson distribution, we used a quasi-Poisson error distribution. To construct the final model, we first ordered the main terms according to decreasing significance in a model with only single terms and no interactions. We then included all interaction terms and removed those that *F* tests revealed did not significantly contribute to the MIC variance observed. The final model (MIC ~ G238S + Exp + E104K + M182T + G238S : Exp) included only significant terms.

MIC Predictions

We used a model first developed by Zimmerman and Rosselet (1977) and expanded upon by Nikaido and Normark (1987) to predict MIC values in our experiments (see also Lakaye *et al.* 1999). Combining Fick's first law of diffusion and

the Michaelis-Menten equation, Zimmerman and Rosselet (1977) derive the following equation for MIC:

$$MIC = C_{inh} + \frac{V_{\max} \cdot C_{inh}}{P \cdot A\left(K_M + C_{inh}\right)} \quad (3)$$

where C_{inh} is the periplamic concentration of β-lactam required to inactivate sufficient numbers of PBPs to inhibit growth, V_{max} is the rate of β-lactamase mediated hydrolysis of β-lactam, P is the permeability coefficient for the specific β-lactam across the outer bacterial membrane (1.8×10^{-5} cm s^{-1} for cefotaxime in *E. coli*), A is the area of membrane per unit of cells (132 cm^2 mg^{-1} for *E. coli*), and K_M is the Michaelis constant for β-lactamase.

Substituting the relationship of $V_{max} = k_{cat}*[E]$ (Nelson and Cox 2000) into equation (3), we rewrite equation 3 as:

$$MIC = C_{inh} + \frac{k_{cat} \cdot [E] \cdot C_{inh}}{P \cdot A\left(K_M + C_{inh}\right)} \quad (4)$$

where k_{cat} is the overall rate constant of the β-lactamase catalyzed β-lactam hydrolysis and $[E]$ is the periplasmic concentration β-lactamase. We inferred the V_{max} against cefotaxime (nmol β-lactam hydrolysed per sec per mg dry weight) that Nikaido and Normark (1987) calculated for cells expressing TEM β-lactamase (contained on plasmid JF701(R_{471a})) based on equation (3). Based on their reported k_{cat} for JF701-encoded TEM, we calculate $[E]$ in Nikaido and Normark (1987)'s experiments to be 0.2508 nmol β-lactam hydrolysed per mg dry weight. To obtain a biologically realistic range of relative enzyme concentrations, we used Nikaido and Normark (1987)'s $[E]$ to anchor the range of relative expression values over which we used equation (4) to predict MIC for each allele.

We used the previously reported k_{cat} and K_M against cefotaxime of four TEM alleles with E at site 104 (Wang *et al*. 2002) along with equation (4) to predict MIC values for each allele over a range of enzyme concentrations (see Figure 4). The absolute enzyme concentrations we measured for each allele were not readily comparable with the enzyme concentration from Nikaido and Normark (1987) (see units of each metric above). Instead, we normalized each value of enzymatic concentration in our experiments by dividing by the lowest enzyme concentration we observed (G238S with no arabinose). This procedure yielded an overall enzymatic expression range of nearly four orders of magnitude with each allele spanning only a part of this range (see Figure 4). Setting the lowest observed enzymatic concentration as 1000-fold lower than the concentration observed by Nikaido and Normark (1987), we predicted MIC values for each allele across its observed relative enzyme concentration (Figure 4, solid lines in each quadrant). For comparison, we then plotted the observed MICs for each allele against their corresponding enzyme concentration (Figure 4, points in each quadrant). Additionally, supplementary Figure S4 depicts MIC prediction for each allele across their entire collective expression range.

Acknowledgments

We thank Pierre Fontanilles, D. Allan Drummond, Christian Landry, Christopher Marx, Sarah Kingan, and Yousif Shamoo for helpful discussions and review of our manuscript. We thank J. Keasling for *E. coli* strain BW27783. K.M.B is supported by an NSF graduate research fellowship. MAD is a Damon Runyon Fellow and was supported by the Damon Runyon Cancer Research Foundation (DRG-1861-05). This work was supported by NIH grant GM079536.

References

Brown KM, Landry CR, Hartl DL, Cavalieri D (2008) Cascading transcriptional effects of a naturally occurring frameshift mutation in *Saccharomyces cerevisiae*. *Molecular Ecology*, **17**, 2985-2997.

Carroll SB (2000) Endless forms: the evolution of gene regulations and morphological diversity. *Cell*, **101**, 577-580.

Carroll SB (2005a) Endless forms most beautiful: the new science of evo devo. W.W. Norton & Co., New York.

Carrol SB (2005b) Evolution at two levels: on genes and form. *PLoS Biology*, **3**, e245.

Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Tod MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferriera S, Wang G, Zheng XG, White TJ, Sninsky JJ, Adams MD, Cargill M (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science*, **302**, 1960-1963.

Dykhuizen DE, Dean AM, Hartl DL (1987) Metabolic Flux and Fitness. *Genetics*, **115**, 25-31.

Ghaemmaghami S, Huh W, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. *Nature*, **425**, 737-741.

Hall BG (1990) Directed evolution of a bacterial operon. *BioEssays*, **12**, 551-558.

Hall BG (1998) Activation of the *bgl* operon by adaptive mutation. *Molecular Biology and Evolution*, **15**, 1-5.

Hall BG, Hartl DL (1974) Regulation of newly evolved enzymes. I. Selection of a novel lactatase regulated by lactose in *Escherichia coli*. *Genetics*, **76**, 391-400.

Hall BG, Hauer B (1993) Acquisition of new metabolic activities by microbial populations. *Methods in Enzymology* **224**, 603-613.

Hall BG, Yokoyama S, Calhoun DH (1983) Role of cryptic genes in microbial evolution. *Molecular Biology and Evolution*, **1**, 109-124.

Hartl, D.L. & Clark A.G. *Principles of Population Genetics* (Sinauer Associates Inc. Publishers, Sunderland, Massachusetts, USA, 2007).

Hartl DL, Dykhuizen DE and Dean AM (1985) Limits of adaptation: the evolution of selective neutrality. *Genetics*, **111**, 655-674.

Hartl DL, Hall BG (1974) Second naturally occurring β–galactosidase in *E. coli*. *Nature*, **248**, 152-153.

Hegeman GD, Rosenberg SL (1970) The evolution of bacterial enzyme systems. *Annual Reviews Microbiology*, **24**, 429-462.

Hoekstra HE, Hirschman RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid replacement contributes to adaptive beach mouse color pattern. *Science*, **313**, 101-104.

Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, **61**, 995-1016.

Khlebnikov A, Datsenko KA, Skaug T, Wanner BL, Keasling JD (2001) Homogeneous expression of the P_{BAD} promoter in *Escherichia coli* by constitutive expression of low-affinity high-capacity AraE transporter. *Microbiology-SGM*, **147**, 3241-3247.

Lakaye B, Dubus A, Lepage S, Groslambert S, Frère JM (1999) When drug inactivation renders the target irrelevant to antibiotic resistance: a case story with β–lactams. *Molecular Microbiology*, **31**, 89-101.

Laraki N, Franceschini N, Rossolini GM, Santucci P, Meunier C, de Pauw E, Amicosante G, Frere JM, Galleni M (1999) Biochemical characterization of the Pseudomonas aeruginosa 101/1477 metallo-beta-lactamase IMP-1 produced by *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, **43**, 902-906.

Lynch VJ, Wagner GP (2008) Resurrecting the role of transcription factor change in developmental evolution. *Evolution*, **62**, 2131-2154.

Mortlock RP, Fossitt DD, Wood WA (1965) A basis for utilization of unnatural pentoses and pentitols by *Aerobacter aerogenes*. *Proceedings of the National Academy of the Sciences*, **54**, 572-579.

Nikaido H, Normark S (1987) Sensitivity of *Escherichia coli* to various β-lactams is determined by the interplay of outer membrane permeability and degradation by periplasmic β-lactamases: a quantitative predicitive treatment. *Molecular Microbiology*, **1**, 29-36.

Nelson DL, Cox MM. *Lehninger Principles of Biochemistry*. (Worth Publishers, New York, NY, 2000).

Olds LC, Sibley E (2003) Lactase persistence DNA variant enhances lactase promoter activity *in vitro*: functional role as a *cis* regulatory element. *Human Molecular Genetics*, **12**, 2333-2340.

Orr HA (2002) The population genetics of adaptation: the adaptation of DNA sequences. *Evolution*, **56**, 1317-1330.

Performance standards for antimicrobial susceptibility testing; Seventeenth informational supplement. (Clinical Laboratory Standards Institute, Wayne, Pennsylvania, USA, 2007).

Stam LF, Laurie CC (1996) Molecular dissection of a major gene effect on a quatitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* **144**, 1559-1564.

Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schulter D, Kingsley DM (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, **428**, 717-723.

Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schulter D, Kingsley DM (2006) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature, **439**, 1014.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, **39**, 31-40.

Wang X, Minasov G, Shoichet BK (2002) Evolution of an Antibiotic Resistance Enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* **320**, 85-95.

Weinreich D.M. Delaney N, Depristo MA, Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111-114 (2006).

Wray GA (2007) The evolutionary significance of *cis*-regulatory mutations. *Nature Reviews Genetics*, **8**, 206-216.

Wu TT, Lin CC, Tanaka S (1968) Mutants of *Aerobacter aerogenes* capable of utilizing xylitol as a novel carbon. *Journal of Bacteriology*, **96**, 447-456.

Zimmerman W, Rosselet A (1977) Function of the outer membrane of *Escherichia coli* as a permeability barrier to beta-lactam antibiotics. *Antimicrobial Agents and Chemotherapy*, **12**, 368-372.

Figure Legends

Figure 1. Contrasting functions that map enzyme activity onto fitness. The dashed line (top) indicates a concave relationship typical for an enzyme in a metabolic pathway. The dotted line (bottom) indicates a convex relationship of the sort predicted for enzyme-mediated antibiotic resistance. The solid line indicates a linear relationship.

Figure 2. Resistance (MIC) versus expression level (arabinose induction) relationships for eight protein-coding alleles. Left panel depicts alleles without glycine at site 238 (Solid line = wild type, dashed line = M182T, dotted line = E104K, dash-dot line = M182T + E104K). Right panel depicts alleles with serine at site 238 (Solid line = G238S, dashed line = G238S + M182T, dotted line = G238S + E104K, dash-dot line = G238S + M182T + E104K). MIC values are $\log_2$ transformed while % arabinose values are $\log_{10}$ transformed.

Figure 3. Comparison of predicted MIC values (solid lines) and observed data (open circles) for four alleles as a function of relative enzyme concentrations ([E]). Relative enzyme concentrations for G238S were adjusted to correct for unusual sensitivity to protein extraction procedures. See Supplementary Figure S4 for MIC predictions for each allele over the entire theoretical range of expression.

Figure 4. These curves illustrate the relative importance of regulatory versus structural mutations in enzyme evolution under concave (top) and convex (bottom) fitness

mappings. Arrows and labels (e.g., “5x”) indicate jumps between curves corresponding to enzymes with increased substrate affinity.

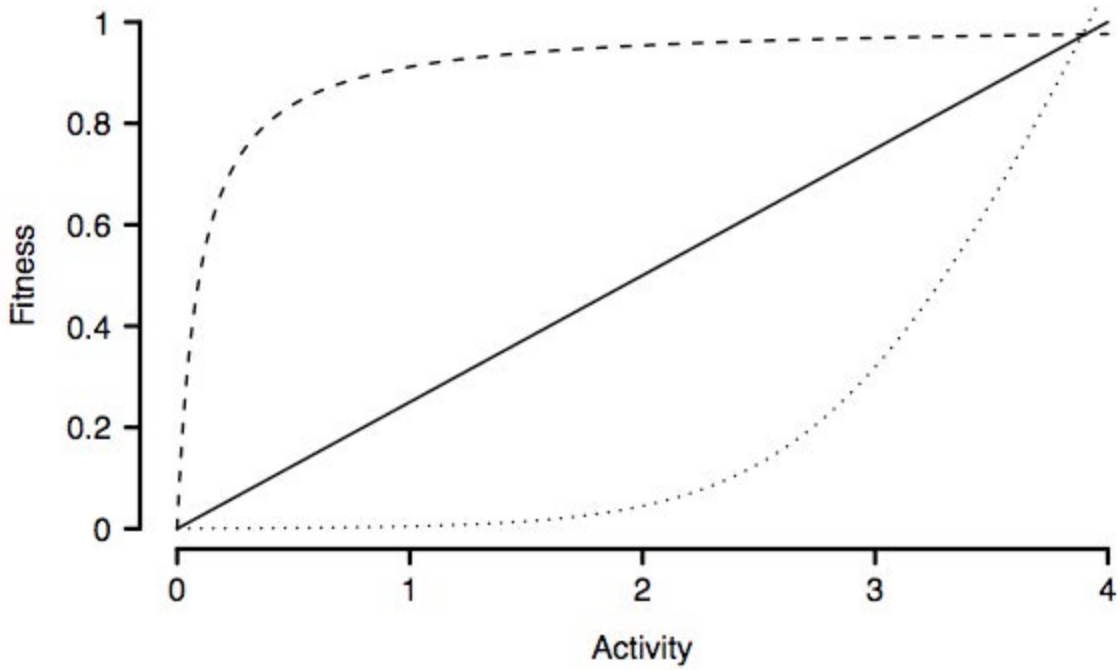


Figure 1.

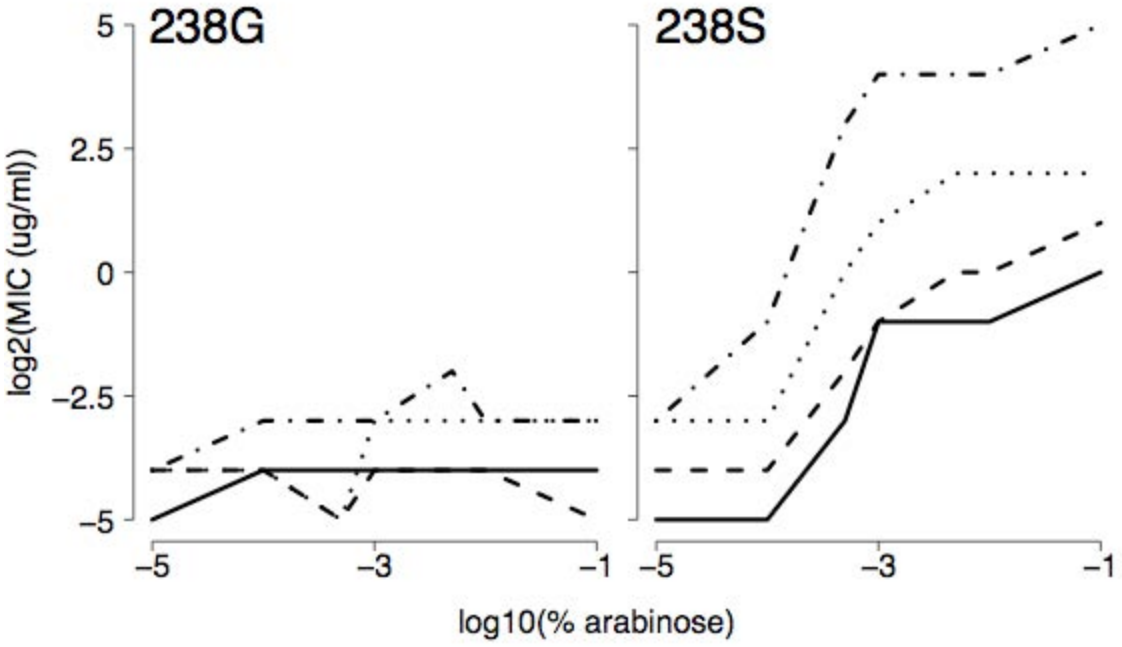


Figure 2.

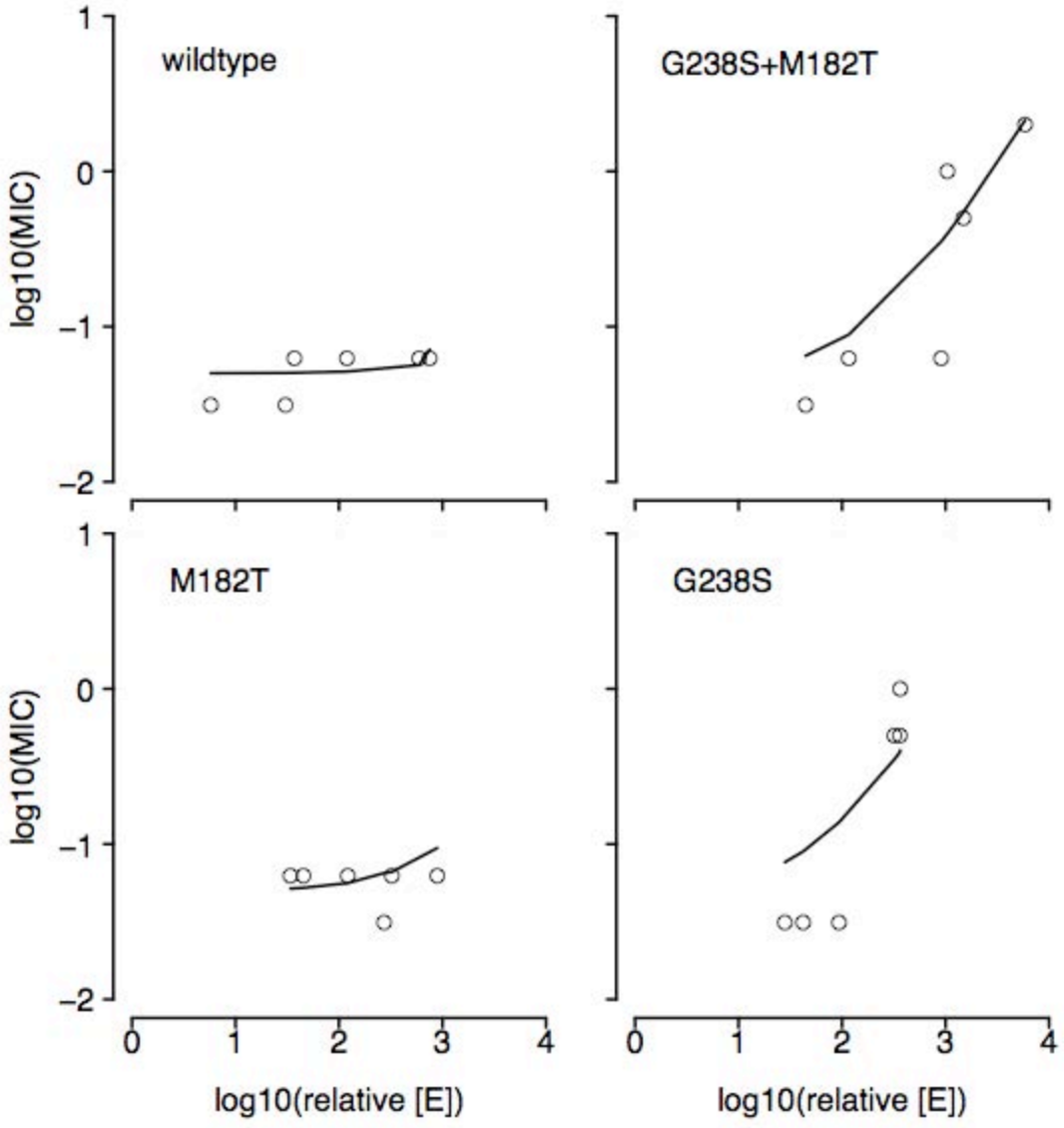

wildtype
G238S+M182T
M182T
G238S
log10(MIC)
log10(MIC)
log10(relative [E])
log10(relative [E])
1
0
−1
−2
0
1
2
3
4

Figure 3.

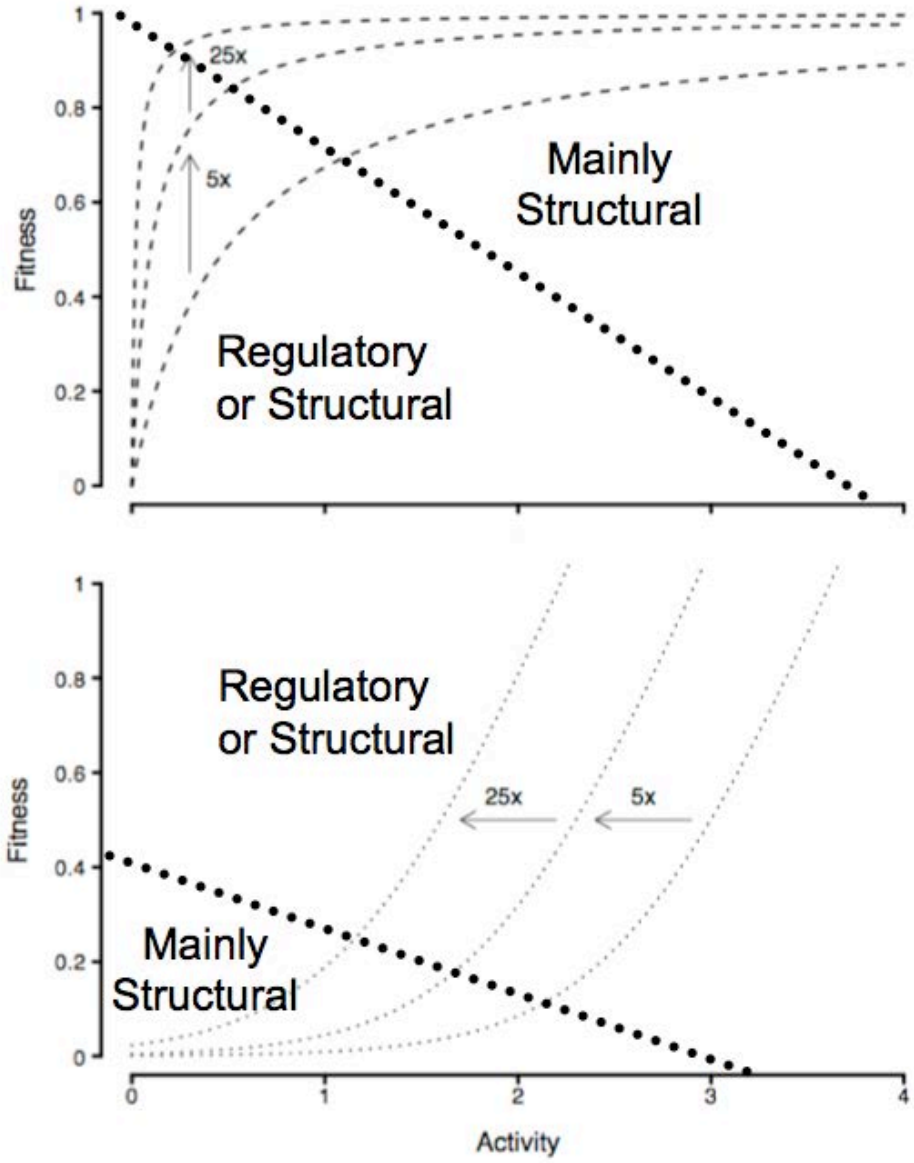
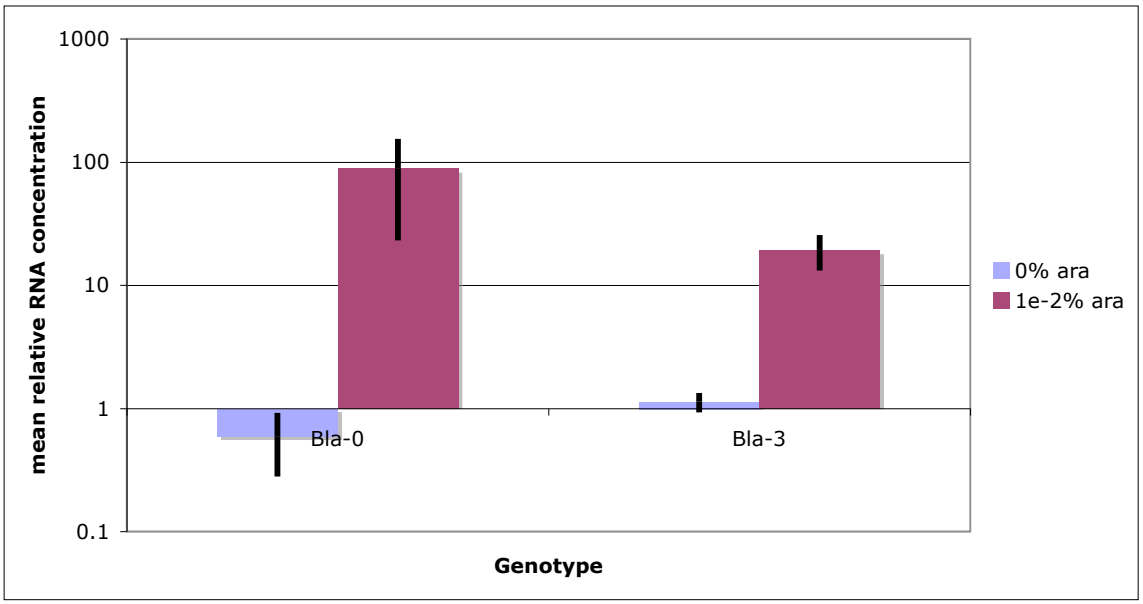
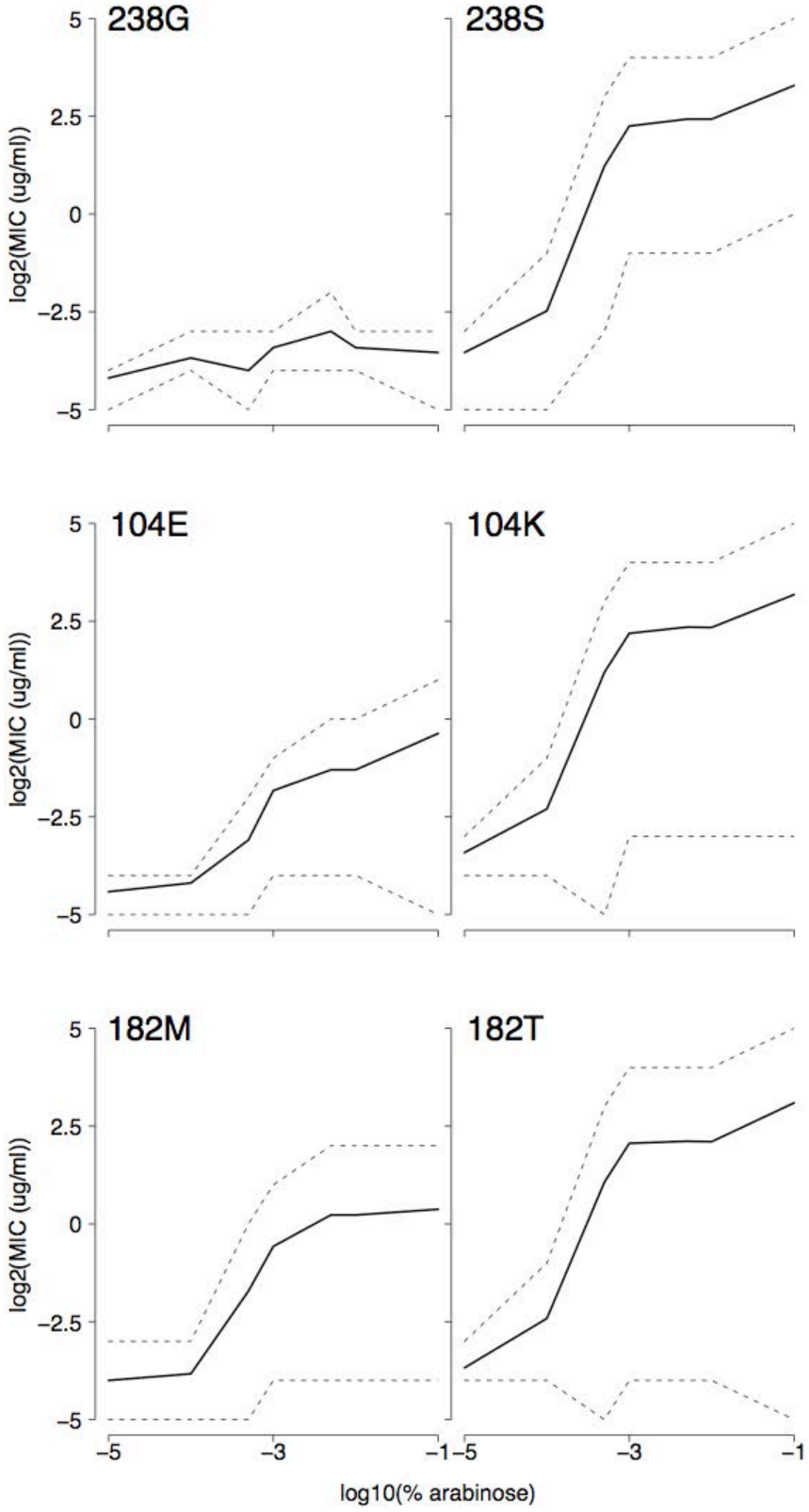

Fitness
1
0.8
0.6
0.4
0.2
0
25x
5x
Mainly
Structural
Regulatory
or Structural
Regulatory
or Structural
25x
5x
Mainly
Structural
0
1
2
3
4
Activity

Figure 4.

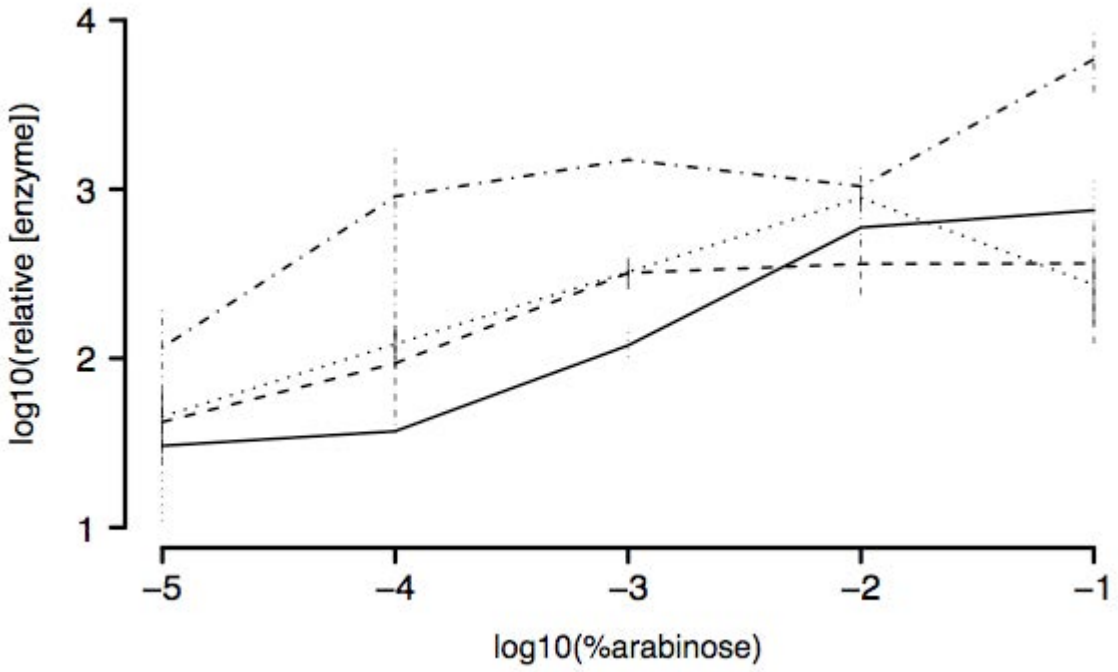
Supplementary Information



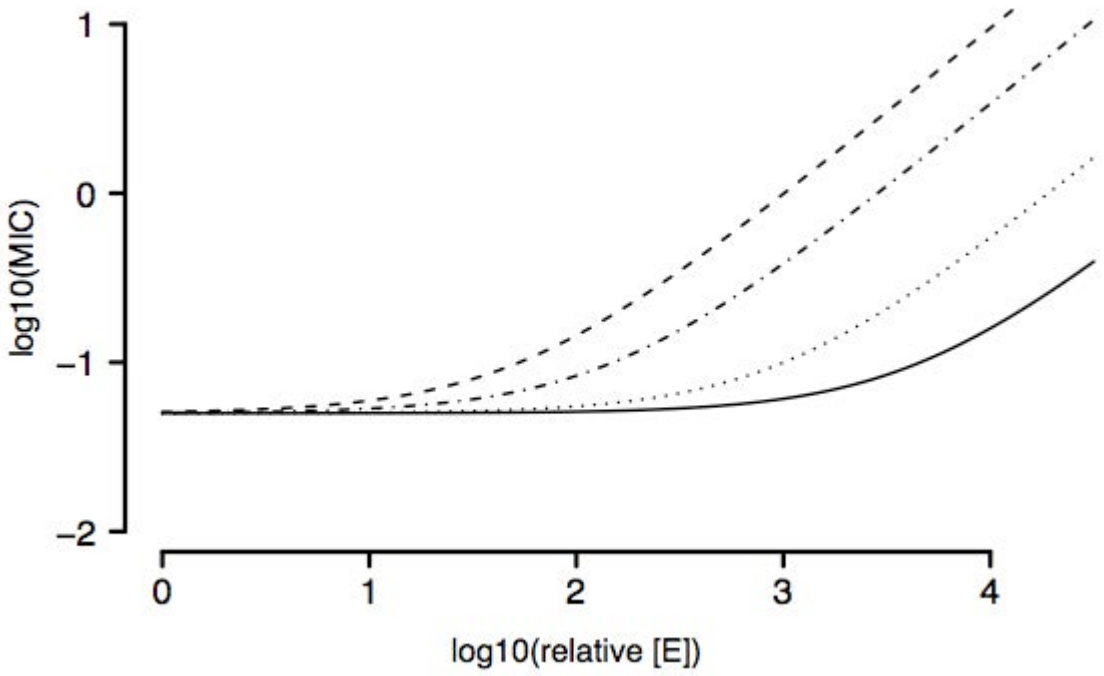
Supplementary Figure S1. Relative expression level of two genotypes at two arabinose concentrations based on TEM β-lactamase RNA concentration. While expression levels at 0% arabinose and 10-2% arabinose are significantly different for both genotypes (heteroscedastic t-test of two samples: Bla-0: p=0.02, Bla-3: p=0.02), expression levels for the two genotypes at the same arabinose concentration are not significantly different (heteroscedastic t-test of two samples: 0%: p=0.19, Bla-3: p=0.22). Statistical analyses based on mean ΔCT values across biological replicates (see Materials and Methods).


238G
238S
104E
104K
182M
182T
log2(MIC (ug/ml))
5
2.5
0
−2.5
−5
−5
−3
−1
log10(% arabinose)

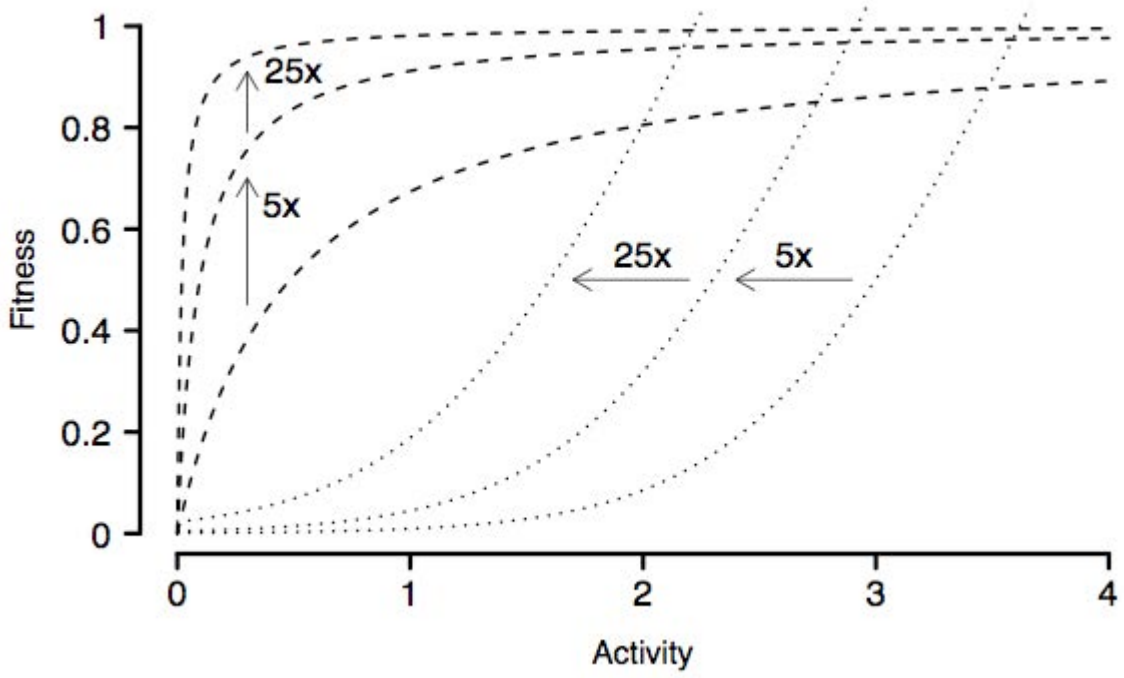
Supplementary Figure S2. Average effect of mutations at amino acid sites 238, 104 and 182. The solid line in each graph represents the average MIC for the four alleles with the amino acid at the site indicated on the graph. Averages from alleles with wild type amino acids at the indicated site are represented on the left while averages from alleles with mutated residues are depicted on the right. Thin dashed lines above and below solid lines indicate the maximum and minimum median MIC value for the 4 alleles averaged at a given arabinose concentration. MIC values are $\log_2$ transformed while % arabinose values are $\log_{10}$ transformed.



Supplementary Figure S3. Relative β-lactamase concentration as a function of arabinose concentration for four alleles. Values were relativized by dividing each absolute expression value by the lowest observed absolute expression value (G238S, 0% arabinose). Points represent the mean of three independent biological replicates. Error bars represent standard error across of biological replicates. To compensate for unusual sensitivity to protein extraction procedures, relative G238S values were multiplied by the average relative expression level of the other three alleles in the absense of arabinose. Solid line = wild type; dashed line = G238S; dotted line = M182T; dash-dot line = G238S + M182T. Complete data set presented in Table S4.



Supplementary Figure S4. MIC predictions for four alleles over the entire expression range. Solid line = wild type; dashed line = G238S; dotted line = M182T; dash-dot line = G238S + M182T.



Supplementary Figure S5. The effect of structural changes versus regulatory changes in enzyme activity on fitness in two different fitness regimes. Dashed lines describe enzymes under a concave fitness-mapping regime while dotted lines describe enzymes under under a convex fitness-mapping regime. Arrows indicate changes in an enzyme's kinetic parameters (k_{cat}/K_M) within a fitness regime. While regulatory changes move an organism along a given curve, structural changes alter the kinetic parameters and change the curve upon which the organism lies.

Supplementary Table S1. TEM β-lactamase alleles and their corresponding Michaelis-Menten and rate constants toward nitrocefin (see Materials and Methods). A "plus" sign indicates the presence of the mutation listed above the column.

Allele[a]	E104K	M182T	G238S	K_M (μM)	sd[b] (K_M)	k_{cat} (1/sec)	sd (k_{cat})
0	-	-	-	50.2	2.7	12800	561
1	-	-	+	11.5	0.8	737	48.8
2	-	+	-	51.9	4.3	16700	982
3	-	+	+	11.4	3.5	203	13.6
4	+	-	-	nd[c]		nd	
5	+	-	+	nd		nd	
6	+	+	-	nd		nd	
7	+	+	+	nd		nd	

[a] For ease of reference, allele designations used in this table are repeated throughout the Supplementary Information.
[b] sd: Standard deviation of the value indicated.
[c] nd: Not determined.

Supplementary Table S2. Median MIC values (μg/ml) for eight alleles at eight arabinose concentrations.

	Arabinose induction							
Allele	0%	10^{-5}%	10^{-4}%	5 x 10^{-4}%	10^{-3}%	5 x 10^{-3}%	10^{-2}%	10^{-1}%
0	0.03	0.03	0.06	0.06	0.06	0.06	0.06	0.06
1	0.03	0.03	0.03	0.13	0.5	0.5	0.5	1
2	0.06	0.06	0.06	0.03	0.06	0.06	0.06	0.03
3	0.03	0.06	0.06	0.25	0.5	1	1	2
4	0.06	0.06	0.06	0.03	0.13	0.13	0.13	0.13
5	0.13	0.13	0.13	1	2	4	4	4
6	0.13	0.06	0.13	0.13	0.13	0.25	0.13	0.13
7	0.25	0.125	0.5	8	16	16	16	32

Supplementary Table S3. Complete MIC data set.

Allele	[Arabinose]	Replicate MIC (μg/ml) values		
0	0%	0.03	0.03	0.03
	10^{-5}%	0.03	0.03	0.03
	10^{-4}%	0.06	0.06	0.06
	5 x 10^{-4}%	0.06	0.06	0.06
	10^{-3}%	0.06	0.06	0.06
	5 x 10^{-3}%	0.06	0.06	0.03
	10^{-2}%	0.06	0.06	0.06
	10^{-1}%	0.06	0.06	0.06
1	0%	0.03	0.03	0.03
	10^{-5}%	0.03	0.03	0.03
	10^{-4}%	0.03	0.03	0.03
	5 x 10^{-4}%	0.25	0.13	0.13
	10^{-3}%	0.5	0.5	0.5
	5 x 10^{-3}%	0.25	0.5	0.5
	10^{-2}%	0.5	0.5	0.5
	10^{-1}%	1	1	1
2	0%	0.06	0.06	0.06
	10^{-5}%	0.06	0.06	0.06
	10^{-4}%	0.06	0.06	0.06
	5 x 10^{-4}%	0.03	0.03	0.03
	10^{-3}%	0.06	0.06	0.06
	5 x 10^{-3}%	0.06	0.06	0.06
	10^{-2}%	0.06	0.06	0.03
	10^{-1}%	0.06	0.03	0.03
3	0%	0.03	0.03	0.06
	10^{-5}%	0.06	0.06	0.06
	10^{-4}%	0.06	0.06	0.06
	5 x 10^{-4}%	0.25	0.125	0.25
	10^{-3}%	0.5	0.5	0.5
	5 x 10^{-3}%	1	2	1
	10^{-2}%	1	2	1
	10^{-1}%	2	4	2

Table S3 (continued).

Allele	[Arabinose]	Replicate MIC (μg/ml) values		
4	0%	0.06	0.13	0.06
	10^{-5}%	0.06	0.13	0.06
	10^{-4}%	0.06	0.13	0.06
	5 x 10^{-4}%	0.06	0.03	0.03
	10^{-3}%	0.13	0.13	0.13
	5 x 10^{-3}%	0.13	0.13	0.13
	10^{-2}%	0.13	0.13	0.13
	10^{-1}%	0.13	0.13	0.13
5	0%	0.13	0.13	0.13
	10^{-5}%	0.13	0.13	0.13
	10^{-4}%	0.13	0.13	0.25
	5 x 10^{-4}%	2	1	1
	10^{-3}%	2	2	2
	5 x 10^{-3}%	4	4	4
	10^{-2}%	4	4	4
	10^{-1}%	4	4	F*
6	0%	0.13	0.13	0.13
	10^{-5}%	0.06	0.06	0.06
	10^{-4}%	0.13	0.13	0.13
	5 x 10^{-4}%	0.13	0.13	0.13
	10^{-3}%	0.13	0.13	0.06
	5 x 10^{-3}%	0.25	0.25	0.25
	10^{-2}%	0.13	0.13	0.13
	10^{-1}%	0.13	0.13	0.13
7	0%	0.25	0.13	0.25
	10^{-5}%	0.13	0.13	0.25
	10^{-4}%	0.25	0.5	0.5
	5 x 10^{-4}%	8	8	8
	10^{-3}%	32	16	8
	5 x 10^{-3}%	16	16	16
	10^{-2}%	16	16	16
	10^{-1}%	64	32	32

* indicates failed MIC test.

Supplementary Table S4. Mean β-lactamase concentration in soluble protein extracts (milliΔOD/(60mg soluble protein extract)) as a function of arabnose induction level. Data used to create Figure 4. Standard error for each value is indicated in the line immediately below the averages (denoted SE).

	Arabinose induction					
Allele	0%	10^{-5}%	10^{-4}%	10^{-3}%	10^{-2}%	10^{-1}%
0	9.5e-6	5.0e-5	6.1e-5	2.0e-4	9.8e-4	1.2e-3
SE	3.3e-6	3.3e-5	6.3e-6	3.8e-5	2.5e-4	6.0e-4
1	1.6e-6	2.5e-6	5.5e-6	1.9e-5	2.1e-5	2.1e-5
SE	6.5e-7	1.2e-6	3.1e-6	7.7e-7	8.1e-6	1.5e-5
2	5.6e-5	7.4e-5	2.0e-4	5.3e-4	1.5e-3	4.5e-4
SE	7.2e-6	3.0e-5	5.7e-5	1.1e-4	2.6e-4	1.9e-4
3	7.3e-5	2.0e-4	1.5e-3	2.4e-3	1.7e-3	9.6e-3
SE	2.2e-5	1.2e-4	1.3e-3	8.5e-5	4.9e-4	4.0e-3